

# #Neuland - Museen und Archive als gleichberechtigte Partner in der GND-Kooperative

Eine Einführung in das Projekt  
„GND für Kulturdaten“ (GND4C)



# Projektpartner



# Motivation

Verlässlichkeit Verlässlichkeit Verlässlichkeit Verlässlichkeit Verlässlichkeit Ve



BRÜCKEN IM NETZWERK DER KULTUR  
UND WISSENSCHAFT

[www.dnb.de/gnd](http://www.dnb.de/gnd)

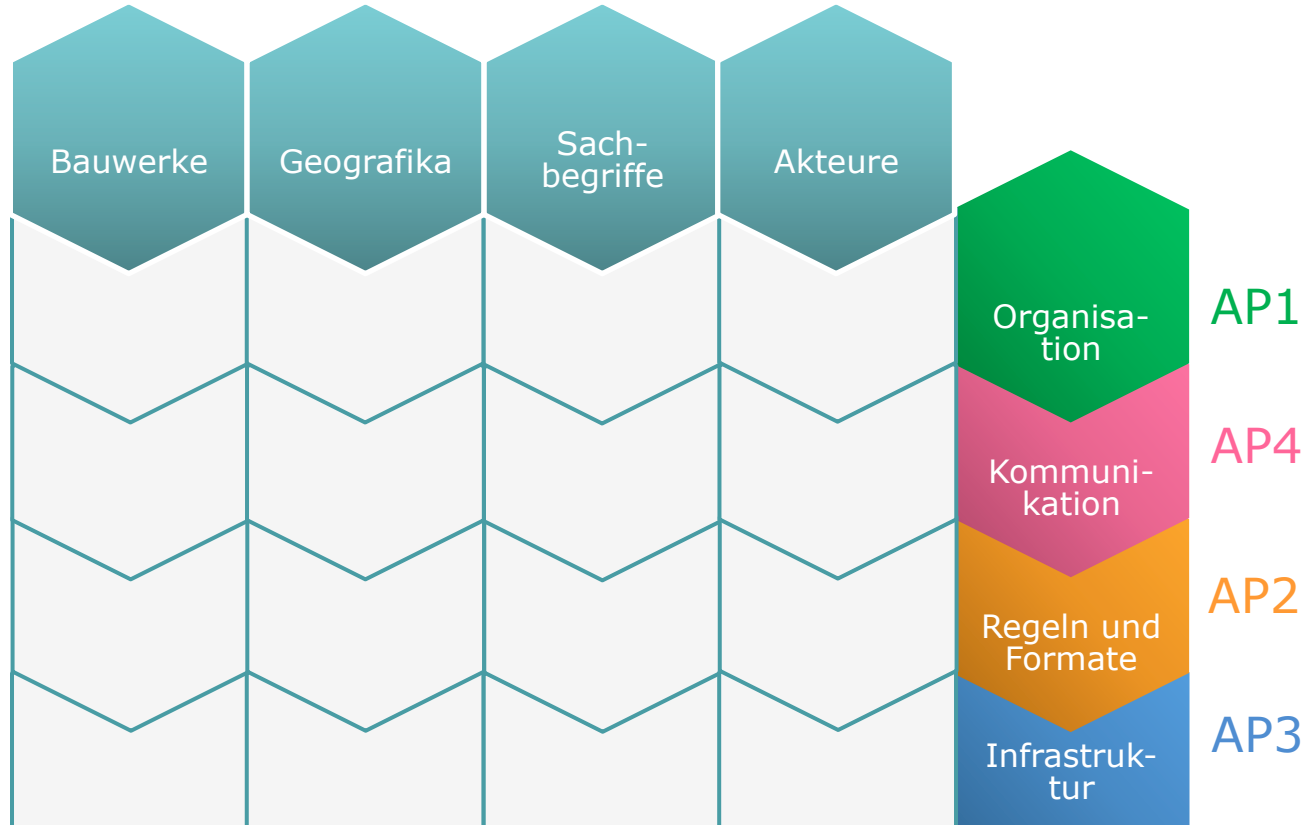
# Laufzeit

- Zwei Projektphasen geplant:
  - Phase 1:
    - Mai 2018 – Okt 2020
    - Konzeption, Vorbereitung, teilweiser Pilotbetrieb
  - Phase 2:
    - Voraussichtlich Nov 2020 – Apr 2023
    - Produktivnahme und Erweiterung

# Aktionsfelder und Projektziele

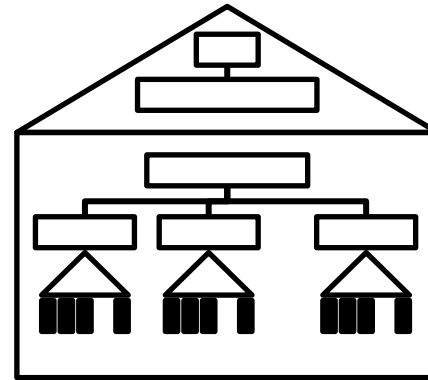


## Fallbeispiele



# AP1 Interdisziplinäre Organisation - Idee & Vorgehen

- Agenturangebote für neue Interessengruppen schaffen
- Strukturellen Bedarf ermitteln
- bestehende Strukturen nutzen:
  - Standardisierungsausschuss (STA)
  - DDB-Fachstellen
- Vereinbarung anpassen



Verlässlichkeit!



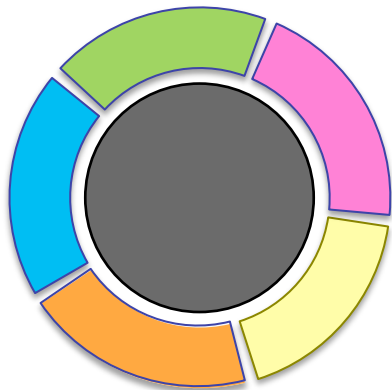
# AP1 Interdisziplinäre Organisation - Bestandsaufnahme

- Bestehenden Agenturen / GND-Zentrale
  - Portfolio heterogen / Teils gute Erweiterbarkeit für neue Interessengruppen gegeben
  - Qualitätssicherung nicht einheitlich und nicht systematisch genug
- neue Interessengruppen
  - Ähnliche Bedarfe aber sehr heterogene Voraussetzungen
  - Zentrale Stellen / Strukturen nur teilweise vorhanden
  - Finanzierung zentraler Dienste große Herausforderung

## AP4 Gemeinsam Arbeiten im GND-Netzwerk - Stand

- Unterschiedliche Kanäle gestartet (u.a. GNDCon)
- GND-Wiki als Dialogplattform im Aufbau
- Verzahnung mit AGs und Expertenteams des STA
- Ausbau des Ticketsystems (JIRA) für kooperative Workflows (QS, Support, Neumitglieder, ...)

## AP2 Regeln und Formate - Idee & Vorgehen



- Anforderungen anhand der Fallbeispiele ermitteln (Gemeinsamkeiten / Unterschiede)
- Umsetzbarkeit mit aktuellen Regeln und Strukturen evaluieren
- Entwicklung kurz- und mittelfristiger Konzepte

## AP2 Stand der Dinge

- Anforderungen aus den Fallbeispielen und AGs des STA
- Kurz bis mittelfristig:
  - **Datenstruktur:** Minimalanforderungen in bestehenden Strukturen umsetzbar
  - **Regeln:** Stehen teilweise im Konflikt mit den Anforderungen / Rahmenbedingungen → Klärungsbedarf
  - **Beispiele:** Regeln zur Bildung von Benennungen, fehlende Beziehungskennzeichen

## AP2 Stand der Dinge

- Mittel- und langfristig:
  - neues Datenmodell (GND2.0)
  - Modulares Konzept für DACH-Regelungen
    - siehe Parallelaktivität RNAB
  - Konzeptentwicklung und Evaluation mittels Wikibase
    - siehe Parallelprojekt „[GND meets Wikibase](#)“

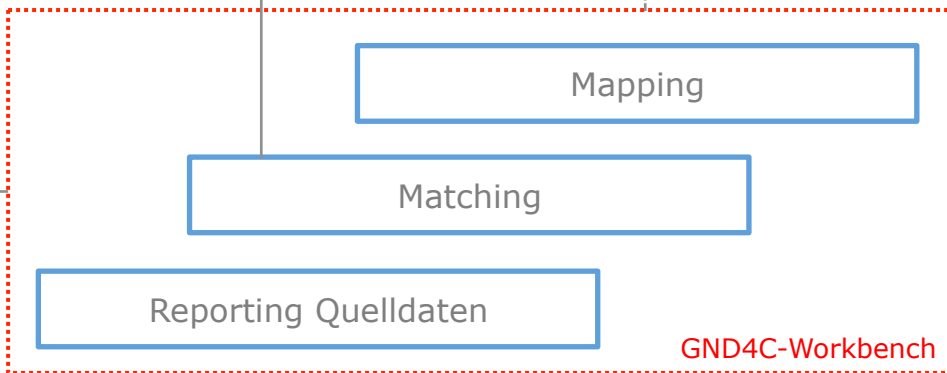
# AP3 Verfahren für die Anbindung neuer Datenquellen – Idee und Vorgehen

- Marktanalyse: existierende Abgleich-Werkzeuge auswerten
- Aufbau eines ProjectLab
- GND4C Workbench
  - Entwicklung eines konfigurierbaren Tools für effiziente Abgleiche eigener Datensets mit der GND
  - Evaluation und Nutzung durch die Fallbeispiele

# AP3 Bedeutung "Matching" und "Mapping"

- Matching: automatischer Abgleich
  - Spezielle Algorithmen und Konfiguration pro Datenset und Entitätstyp
  - Ergebnis: 0 bis n Matches
  - Matches sind Kandidaten für den Mapping-Prozess
- Mapping: intellektueller Abgleich
  - semantisch übereinstimmende Begriffe/Entitäten festlegen
  - Art der Übereinstimmung kann qualifiziert werden:
    - z.B. exakt oder annähernd übereinstimmend

**APIs**  
**Normdatensysteme:**  
u.a. GND4C-RDF (GND),  
lobid (GND), Wikibase  
(GND), AAT, Wikidata,  
GeoNames, OSM, ...



Normalisierte Daten

Quelldaten Fallbeispiele

Personen, Geographika,  
Bauwerke, Sachbegriffe

Erfassungssysteme Fallbeispiele



Infrastruktur



# AP3 Stand der Dinge: ProjektLab

- GND4C RDF-API. SPARQL-Endpoint:  
<http://gnd4c.digicult-verbund.de/sparql.html>
- Technische Analyse der Lieferdaten aus den Fallbeispielen  
→ Normalisierter Datenspeicher
- Tests Schnittstellen auf Normdaten mit Matchingexperimenten
  - u.a. Lobid, DNB-SRU, GND4C RDF-API, Wikibase/Wikidata, OpenStreetMap, AAT deutsch, AAT Getty
- Analyse von Wikibase  
→ u.a. Installation von Testinstanzen für semantische Experimente
- GND4C-Oberfläche (Workbench)
  - Reporting der Quelldaten zur Unterstützung der Qualitätssicherung
  - Semi-automatischer Abgleich (Matching + Mapping) der Lieferdaten mit der GND als Hauptreferenz
  - Konfigurierbare Abfragen/Algorithmen

# AP3 Stand der Dinge: Workbench

- Version 1 mit Draft 1 ab ca. Mitte Mai 2019
  - Zunächst Focus Personen (Entität Person, Geographika, Sachbegriff)
  - Reporting Quelldaten zur Unterstützung der Qualitätssicherung
    - Häufigkeitsanalyse Tätigkeit/Beruf bei Personen
    - Auswertung Mapping Personen mit/ohne GND-URI
    - Auswertung Geburts-/Sterbeort Personen
    - Auswertung über ISO-8601-konforme Datierungsangaben
    - ...
  - Mehrstufige Matchingdurchläufe werden als Matchingereignisse gespeichert
- In Vorbereitung:
  - Paralleles Matching über unterschiedliche Normvokabulare
  - Pro Entitätstyp jeweils eigene Suchstrategien
  - Erweiterte Suche über Fuzzy-Search, Stemming, Levenshtein-Distanz
  - Regeln und Berechnung für Konfidenzwerte. Darstellung über "Farbampel"
  - Intellektuelles Clustern von Quelldaten unter Berücksichtigung der Matchingereignisse
  - Intellektueller Abgleich (Mapping) auf der Basis der Matching-Vorschläge

# Experimentelles Abgleichverfahren: Version 0 (Nov. 2018-Jan. 2019)

Willkommen | Arbeitsbereich | Einstellungen

Suche | Anmerkungen Suche | Abgleich

### Experimentelles Abgleichverfahren GND4C (Version 0)

**Quelldaten:** imdas pro / Städtische Museen Freiburg | Version 0

**Ergebnis Limit:** alle | **Entität:** Person | **Zielvokabular:** GND | **Zugriffsart:** lobid |

Nr	Hits	surname	forename	gender	dateOfBirth	placeOfBirth	dateOfDeath	placeOfDeath	associatedPlace	profession OrOccupation	biographicalOrHistoricalInform
1	1	Aldegrevér	Heinrich	M	(um) 1502				Geburtsort: Paderborn; Todesort: Paderborn	Goldschmied; Kupferstecher; Maler; Zeichner	Quelle zum Thieme-Becker-Namer ThB1, 1907
2	2	Ammann	Jost	M	(vor) 13.06.1539		(vor) 1605-1691		Todesort: Nürnberg; Geburtsort: Zürich	Künstler/in; Maler; Radierer; Zeichner	Quelle zum Thieme-Becker-Namer ThB1, 1907
3	2	Arnout	Jean-Baptiste	M	16.174		1865			Entwerfer; Graphiker; Künstler/in; Lithograph; Maler; Zeichner	Quelle zum Thieme-Becker-Namer ThB2, 1908
4	2	Arntz	Gerd		11.11.1900		08.12.1988			Graphiker; Illustrator; Künstler/in; Maler	
5	2	Arthur, Großbritannien, Prinz									
6	5	Augustin le Grand								Künstler/in; Kupferstecher; Stecher	

1 lobid DP [124841260](#) Bruzen de La Martinière, Antoine Augustin Polyhistor 1662-1746

2 lobid DP [131614878](#) Grangé, Jean Augustin Verleger, Drucker, Buchdrucker, Buchhändler 1712-1791

3 lobid DP [1157717101](#) LeGrand de Laleu, Louis Augustin -

4 lobid DP [1018466029](#) Legrand, Augustin -

5 lobid DP [100506194](#) Legrand, Augustin Verleger, Zeichner, Kupferstecher 1765-1815

# Einige Weiterentwicklungen für Version 1 beim Matching

Willkommen | Arbeitsbereich | Einstellungen

Suche | Anmerkungen Suche | Abgleich

### Experimentelles Abgleichverfahren GND4C (Version 0)

**Quelldaten:** imdas pro / Städtische Museen Freiburg | Version 0

**Ergebnis** | **Entität:** | **Zielvokabular:** GND | **Zugriffsart:** lobid | Go

Nr	Hits	surname	forename	gender	dateOfBirth	placeOfBirth	dateOfDeath	placeOfDeath	associatedPlace	profession OrOccupation	biographicalOrHistoricalInform
1	1	Aldegrever	Heinrich	M	(um) 1502				Geburtsort: Paderborn; Todesort: Soest	Goldschmied; Künstler/in; Kupferstecher; Maler; Zeichner	Quelle zum Thieme-Becker-Namer ThB1, 1907
2	2	Ammar								Künstler/in; Maler;	Quelle zum Thieme-Becker-Namer ThB1, 1907
3	2	Arnout								Radierer; Zeichner	
4	2	Arntz								Entwerfer; Graphiker; Künstler/in; Lithograph; Maler; Zeichner	Quelle zum Thieme-Becker-Namer ThB2, 1908
5	2	Arthur, Großbritannien, Prinz								Graphiker; Illustrator; Künstler/in; Maler	
6	5	Augustin le Grand								Künstler/in; Kupferstecher; Stecher	

1 lobid DP [124841260](#) Bruzen de La Martinière, Antoine Augustin Polyhistor 1662-1746  
2 lobid DP [131614878](#) Grangé, Jean Augustin Verleger, Drucker, Buchdrucker, Buchhändler 1712-1791  
3 lobid DP [1157717101](#) LeGrand de Laleu, Louis Augustin -  
4 lobid DP [1018466029](#) Legrand, Augustin -  
5 lobid DP [100506194](#) Legrand, Augustin Verleger, Zeichner, Kupferstecher 1765-1815

Paralleles Matching über unterschiedliche Normvokabulare/Dienste.  
Bei Personen:  
Lobid, GND4C RDF-API, Wikidata

Neues Basisdesign.  
Tests mit Klickpfaden für Eingabe in den Formularen.

# Einige Weiterentwicklungen für Version 1 beim Matching

Willkommen Arbeitsbereich Einstellungen

Suche Anmerkungen Suche Abgleich

Experimentelles Abgleichverfahren GND4C (Version 0)

Quelldaten: imdas pro / Städtische Museen Freiburg Version 0

Ergebnis Limit: alle

Entität: Person

Zielvokabular: GND

Zugriffsart: lobid

Go

Nr Hits	surname	forename	gender	dateOfBirth	placeOfBirth	dateOfDeath	placeOfDeath	associatedPlace	profession OrOccupation	biographicalOrHistoricalInform
1	Aldegrevier	Heinrich	M	(um) 1502				Geburtsort:	Goldschmied;	Quelle zum Thieme-Becker-Namer

Pro Entitätstyp jeweils eigene Suchstrategie.  
Z.B. Personen:  
1. Suche nach Name, Geburtsdatum, Geburtsort, Sterbedatum, Sterbeort, Beruf/Tätigkeit  
wenn kein Treffer dann  
2. Suche nach Name, Geburtsdatum, Geburtsort  
wenn kein Treffer dann  
3. Suche nach Name, Geburtsdatum  
wenn kein Treffer dann  
4. Suche nach Name  
wenn kein Treffer dann  
5. Unscharfe Suche nach Name  
...  
Regeln werden iterativ mit AP2 angepasst

5 lobid: DP 100500194 Legrand, Augustin verleger, Zeichner, Kupferstecher 1765-1815

Speichern von Matching-durchläufen

# Einige Weiterentwicklungen für Version 1 beim Matching

Willkommen | Arbeitsbereich | Einstellungen

Suche | Anmerkungen Suche | Abgleich

### Experimentelles Abgleichverfahren GND4C (Version 0)

**Quelldaten:** imdas pro / Städtische Museen Freiburg | Version 0

**Ergebnis Limit:** alle | **Entität:** Person | **Zielvokabular:** GND | **Zugriffsart:** lobid |

Nr	Hits	surname	forename	gender	dateOfBirth	placeOfBirth	dateOfDeath	placeOfDeath	associatedPlace	profession	biographicalOrHistoricalInform
1	1	Aldegrevor	Heinrich	M	(um) 1502				Geburtsort: Paderborn; Todesort: Soest	Goldschmied; Künstler/in; Kupferstecher; Maler; Zeichner	Quelle zum Thieme-Becker-Namer ThB1, 1907
2	2	Ammann	Jost	M	(vor) 13.06.1539		(vor) 17.03.1591	Todesort: Nürnberg;		Künstler/in; Maler; Radierer; Zeichner	Quelle zum Thieme-Becker-Namer ThB1, 1907
3	2	Arnout								Entwerfer; Graphiker; Künstler/in; Lithograph; Maler; Zeichner	Quelle zum Thieme-Becker-Namer ThB2, 1908
4	2	Arntz								Graphiker; Illustrator; Künstler/in; Maler	
5	2	Arthur, Großbritannien Prinz									
6	5	Augustin le Grand								Künstler/in; Kupferstecher; Stecher	

1 lobid DP [124841260](#) Bruzen de La Martinière, Antoine Augustin Polyhistor 1662-1746  
2 lobid DP [131614878](#) Grangé, Jean Augustin Verleger, Drucker, Buchdrucker, Buchhändler 1712-1791  
3 lobid DP [1157717101](#) LeGrand de Laleu, Louis Augustin -  
4 lobid DP [1018466029](#) Legrand, Augustin -  
5 lobid DP [100506194](#) Legrand, Augustin Verleger, Zeichner, Kupferstecher 1765-1815

**Unschärfe Zeitintervalle**  
Regeln für *um*, *vor*, *nicht vor*, *nach*, *ca.*, ... werden über die Konfiguration verwaltet.  
Beispiel:  
- (um) 1502 entspricht 1499-1505  
weitere Beispiele:  
- Geburtstag: 1502 oder 1505 oder 1510  
- Geburtstag: 1952 oder 07.07.1953

Vielen Dank!